



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|--|-------------|----------------------------|---------------------|------------------|
| 10/686,964 | 10/15/2003 | Srinivasan Balasubramanian | ARC920030083US1 | 8831 |
| 28342 | 7590 | 12/27/2006 | EXAMINER | |
| SAMUEL A. KASSATLY LAW OFFICE 20690 VIEW OAKS WAY SAN JOSE, CA 95120 | | | DWIVEDI, MAHESH H | |
| | | | ART UNIT | PAPER NUMBER |
| | | | 2168 | |
| SHORTENED STATUTORY PERIOD OF RESPONSE | | MAIL DATE | DELIVERY MODE | |
| 3 MONTHS | | 12/27/2006 | PAPER | |

Please find below and/or attached an Office communication concerning this application or proceeding.

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

| | | |
|------------------------------|-------------------|------------------------|
| Office Action Summary | Application No. | Applicant(s) |
| | 10/686,964 | BALASUBRAMANIAN ET AL. |
| | Examiner | Art Unit |
| | Mahesh H. Dwivedi | 2168 |

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) Responsive to communication(s) filed on 15 October 2003.
- 2a) This action is FINAL. 2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) Claim(s) 1-20 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) Claim(s) _____ is/are allowed.
- 6) Claim(s) 1-20 is/are rejected.
- 7) Claim(s) _____ is/are objected to.
- 8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on 10/15/2003 is/are: a) accepted or b) objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) All b) Some * c) None of:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) Notice of References Cited (PTO-892)
- 2) Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date 10/15/2003.
- 4) Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____.
- 5) Notice of Informal Patent Application
- 6) Other: _____.

DETAILED ACTION

Information Disclosure Statement

1. The information disclosure statement (IDS) submitted on 10/15/2003 has been received, entered into the record, and considered. The submission is in compliance with the provisions of 37 CFR 1.97. Accordingly, the information disclosure statement is being considered by the examiner.

Claim Rejections - 35 USC § 112

2. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

3. Claim 11 recites the limitation "the web space rules" in Page 24. There is insufficient antecedent basis for this limitation in the claim.

Claim 19 recites the limitation "system" in Page 26. There is insufficient antecedent basis for this limitation in the claim.

Claim 20 recites the limitation "system" in Page 26. There is insufficient antecedent basis for this limitation in the claim.

Claim Rejections - 35 USC § 101

5. 35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

6. Claims 14-17, and 19-20 are rejected under 35 U.S.C. 101 as being directed non-statutory subject matter. The language of the claim raises a question as to whether the claim is directed merely to an abstract idea that is not tied to technological art, environment or machine which would result in a practical application producing a concrete useful, and tangible result to form the basis of statutory subject matter under 35 U.S.C. 101.

Software or program can be stored on a medium and/or executed by a computer. In other words, software must be computer readable. The use of computer is not evident in these claims.

7. **For your reference, below is a section from MPEP 2105 :**

(a) Functional Descriptive Material: "Data Structures" Representing Descriptive Material Per Se or Computer Programs Representing Computer Listings Per Se

Data structures not claimed as embodied in computer-readable media are descriptive material per se and are not statutory because they are not capable of causing functional change in the computer. See, e.g., Warmerdam, 33 F.3d at 1361, 31 USPQ2d at 1760 (claim to a data structure per se held nonstatutory). Such claimed data structures do not define any structural and functional interrelationships between the data structure and other claimed aspects of the invention which permit the data structure's functionality to be realized. In contrast, a claimed computer-readable medium encoded with a data structure defines structural and functional interrelationships between the data structure and the computer software and hardware components which permit the data structure's functionality to be realized, and is thus statutory. Similarly, computer programs claimed as computer listings per se, i.e., the descriptions or expressions of the programs, are not physical "things." They are neither computer components nor statutory processes, as they are not "acts" being performed. Such claimed computer programs do not define any structural and functional interrelationships between the computer program and other claimed elements of a computer which permit the computer program's functionality to be realized. In contrast, a claimed computer-readable medium encoded with a computer program is a computer element which defines structural and functional interrelationships between the computer program and the rest of the computer which permit the computer program's functionality to be realized, and is thus statutory. Accordingly, it is important to distinguish claims that define descriptive material per se from claims that define statutory inventions. Computer programs are often recited as part of a claim. Office personnel should determine whether the computer program is being claimed as part of a otherwise

statutory manufacture or machine. In such a case, the claim remains statutory irrespective of the fact that a computer program is included in the claim. The same result occurs when a computer program is used in a computerized process where the computer executes the instructions set forth in the computer program. Only when the claimed invention taken as a whole is directed to a mere program listing, i.e., to only its description or expression, is it descriptive material per se and hence nonstatutory.

Since a computer program is merely a set of instructions capable of being executed by a computer, the computer program itself is not a process and Office personnel should treat a claim for a computer program, without the computer-readable medium needed to realize the computer program's functionality, as nonstatutory functional descriptive material. When a computer program is claimed in a process where the computer is executing the computer program's instructions, Office personnel should treat the claim as a process claim. See paragraph IV.B.2(b), below. When a computer program is recited in conjunction with a physical structure, such as a computer memory, Office personnel should treat the claim as a product claim.

Claim Rejections - 35 USC § 102

8. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

9. Claims 1-9, and 14-20 are rejected under 35 U.S.C. 102(b) as being anticipated by Chakrabarti et al. (U.S. Patent 6,418,433).

10. Regarding claim 1, **Chankrabarti** teaches a method comprising:

- A) selectively prioritizing the documents to crawl based on a set of rules (Column 8, lines 2-30);
- B) fetching prioritized documents from the network (Column 5, lines 40-46);
- C) for each fetched document, determining whether the fetched document is relevant to any of the multiple focus topics (Column 2, lines 56-60, Column 3, lines 51-55, Column 4, lines 61-65, Column 10, lines 18-43);
- D) crawling the fetched document that matches any of the multiple focus topics (Column 2, lines 56-60, Column 3, lines 51-55, Column 4, lines 61-65, Column 10, lines 35-43); and
- E) further crawling out-links on the fetched document based on an assumption that if the fetched document is of interest, the out-links are also of interest (Column 2, lines 56-60, Column 3, lines 51-55, Column 4, lines 61-65, Column 10, lines 35-43).

The examiner notes that **Chankrabarti** teaches “**selectively prioritizing the documents to crawl based on a set of rules**” as “The priority and relevance fields permit two types of crawl policies, i.e., the above-mentioned “soft” and “hard” crawl policies (Column 8, lines 8-11). The examiner further notes that **Chankrabarti** teaches “**fetching prioritized documents from the network**” as “the Web page table 32 includes a priority field 42 that represents how often the Web page is to be revisited by the crawler 14” (Column 5, lines 41-42). The examiner further notes that **Chankrabarti** teaches “**for each fetched document, determining whether the fetched document is relevant to any of the multiple focus topics**” as “The topic analyzer 28 compares

the content of a Web page with a predefined topic or topics and generates a response representative of how relevant the Web page is" (Column 4, lines 61-65), "When the process determines that the page under test is not relevant to the predefined topic" (Column 10, lines 18-19), and "If the page under test is determined to be relevant to the topic" (Column 10, lines 35-36). The examiner further notes that **Chankrabarti** teaches "**crawling the fetched document that matches any of the multiple focus topics**" as "If the page under test is determined to be relevant to the topic, however, the process moves to block 110, wherein entries are generated for the link table 34 for all outlinks of the page" (Column 10, lines 35-39). The examiner further notes that **Chankrabarti** teaches "**further crawling out-links on the fetched document based on an assumption that if the fetched document is of interest, the out-links are also of interest**" as "If the page under test is determined to be relevant to the topic, however, the process moves to block 110, wherein entries are generated for the link table 34 for all outlinks of the page" (Column 10, lines 35-39).

Chankrabarti does not explicitly teach:

E) selectively prioritizing the documents to crawl based on a set of rules.

Meyerzon, however, teaches "**selectively prioritizing the documents to crawl based on a set of rules**" as "Active plug-ins may be used...a gatherer project may be created that seeks to index all Web documents found during a Web" (Column 11, lines 13-20).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching

Meyerzon's would have allowed **Chankrabarti's** to provide a method to minimize or eliminate time-consuming and error-prone acknowledgement by the patron or delivery personnel to achieve an acceptable degree of efficiency, as noted by **Meyerzon** (Column 11, lines 2-8).

Regarding claim 2, **Chankrabarti** teaches a method comprising:

A) seeding a plurality of seed uniform resource locator strings to start the collaborative focused crawling of the documents (Chakrabarti, Column 5, lines 61-67-Column 6, lines 1-15).

The examiner notes that **Chankrabarti** teaches “**seeding a plurality of seed uniform resource locator strings to start the collaborative focused crawling of the documents**” as “It is to be understood that information pertaining to a “seed” set of Web pages is initially stored in the Web page table 32. The seed set can be gathered from, e.g., the temporary Internet file directories of the employees of a company or from some other group that can be expected to have shared interests... Thus, the seed set does not define a comprehensive, universal set of all topics on the Web, but rather a relatively narrow topic or range of topics that are of interest to the particular source” (Column 5, lines 61-67-Column 6, lines 1-4).

Regarding claim 3, **Chankrabarti** teaches a method comprising:

A) crawling the seed uniform resource locator strings (Chakrabarti, Column 6, lines 61-67-Column 7, lines 1-2, Column 10, lines 44-64).

The examiner notes that **Chankrabarti** teaches “**crawling the seed uniform resource locator strings**” as “starting with the seed set the URL of each page is selected” (Column 6, lines 61-62) and “the current page is classified to its topics, using the topic analyzer 28 (FIG. 1), and then the page is evaluated for relevancy to the predefined topic at the decision diamond 116...when the page is a “good” page the logic expands the outlinks of the page” (Column 10, lines 45-51).

Regarding claim 4, **Chankrabarti** teaches a method comprising:

A) writing a plurality of resulting uniform resource locator strings obtained by crawling the seed uniform resource locator strings (Column 10, lines 35-43, 51-64).

The examiner notes that **Chankrabarti** teaches “**writing a plurality of resulting uniform resource locator strings obtained by crawling the seed uniform resource locator strings**” as “If the page under test is determined to be relevant to the topic, however, the process moves to block 110, wherein entries are generated for the link table 34 for all outlinks of the page” (Column 10, lines 35-38).

Regarding claim 5, **Chankrabarti** teaches a method comprising:

A) a foreman function for reading a plurality of contents of the resulting uniform resource locator strings (Column 10, lines 4-10, 51-64)

The examiner notes that **Chankrabarti** teaches “**a foreman function for reading a plurality of contents of the resulting uniform resource locator strings**” as “If the checksum comparison at decision diamond 100 indicates that new data is

begin considered, however, the logic proceeds to block 102 to tokenize the Web page" (Column 10, lines 4-6).

Regarding claim 6, **Chankrabarti** teaches a method comprising:

A) the foreman function passing the contents of the resulting uniform resource locator strings to a miner (Column 10, lines 10-17, 51-64).

The examiner notes that **Chankrabarti** teaches "**a foreman function for reading a plurality of contents of the resulting uniform resource locator strings**" as "Then , the page is classified at block 104 using the topic analyzer or classifier 28" (Column 10, lines 10-11).

Regarding claim 7, **Chankrabarti** teaches a method comprising:

A) the miner instructing a fetcher to crawl a plurality of out-links on a document of the resulting resource locator string when the contents of the resulting resource locator string match a focus topic of the miner (Column 10, lines 35-43, 51-64).

The examiner notes that **Chankrabarti** teaches "**the miner instructing a fetcher to crawl a plurality of out-links on a document of the resulting resource locator string when the contents of the resulting resource locator string match a focus topic of the miner**" as "If the page under test is determined to be relevant to the topic, however, the process moves to block 110, wherein entries are generated for the link table 34 for all outlinks of the page" (Column 10, lines 35-38).

Regarding claim 8, **Chankrabarti** teaches a method comprising:

A) the miner ignoring resulting resource locator string when the contents of the resulting resource locator string do not match the focus of the miner (Column 10, lines 18-34).

The examiner notes that **Chankrabarti** teaches “**the miner instructing a fetcher to crawl a plurality of out-links on a document of the resulting resource locator string when the contents of the resulting resource locator string match a focus topic of the miner**” as “When the process determines that the page under test is not relevant to the predefined topic, the process moves to block 108 to update the Web page table 32...the outlinks of the page under test are not entered into the link table” (Column 10, lines 18-24).

Regarding claim 9, **Chankrabarti** teaches a method comprising:

A) the miner managing a plurality of focus topics (Column 2, lines 56-60, Column 3, lines 51-55, Column 4, lines 61-65).

The examiner notes that **Chankrabarti** teaches “**the miner managing a plurality of focus topics**” as “The topic analyzer 28 compares the content of a Web page with a predefined topic or topics and generates a response representative of how relevant the Web page is” (Column 4, lines 61-65).

Regarding claim 14, **Chankrabarti** teaches a computer program product comprising:

- A) a first set of instruction codes for selectively prioritizing the documents to crawl based on a set of rules (Column 8, lines 2-30);
- B) a second set of instruction codes for fetching prioritized documents from the network (Column 5, lines 40-46);
- C) for each fetched document, a third set of instruction codes determines whether the fetched document is relevant to any of the multiple focus topics (Column 2, lines 56-60, Column 3, lines 51-55, Column 4, lines 61-65, Column 10, lines 18-43);
- D) a fourth set of instruction codes for crawling the fetched document that matches any of the multiple focus topics (Column 2, lines 56-60, Column 3, lines 51-55, Column 4, lines 61-65, Column 10, lines 35-43); and
- E) wherein the fourth set of instruction codes further crawls out-links on the fetched document based on an assumption that if the fetched document is of interest, the out-links are also of interest (Column 2, lines 56-60, Column 3, lines 51-55, Column 4, lines 61-65, Column 10, lines 35-43).

The examiner notes that **Chankrabarti** teaches “**a first set of instruction codes for selectively prioritizing the documents to crawl based on a set of rules**” as “The priority and relevance fields permit two types of crawl policies, i.e., the above-mentioned “soft” and “hard” crawl policies (Column 8, lines 8-11). The examiner further notes that **Chankrabarti** teaches “**a second set of instruction codes for fetching prioritized documents from the network**” as “the Web page table 32 includes a priority field 42 that represents how often the Web page is to be revisited by the crawler 14” (Column 5, lines 41-42). The examiner further notes that **Chankrabarti** teaches

“for each fetched document, a third set of instruction codes determines whether the fetched document is relevant to any of the multiple focus topics” as “The topic analyzer 28 compares the content of a Web page with a predefined topic or topics and generates a response representative of how relevant the Web page is” (Column 4, lines 61-65), “When the process determines that the page under test is not relevant to the predefined topic” (Column 10, lines 18-19), and “If the page under test is determined to be relevant to the topic” (Column 10, lines 35-36). The examiner further notes that **Chankrabarti** teaches **“a fourth set of instruction codes for crawling the fetched document that matches any of the multiple focus topics”** as “If the page under test is determined to be relevant to the topic, however, the process moves to block 110, wherein entries are generated for the link table 34 for all outlinks of the page” (Column 10, lines 35-39). The examiner further notes that **Chankrabarti** teaches **“wherein the fourth set of instruction codes further crawls out-links on the fetched document based on an assumption that if the fetched document is of interest, the out-links are also of interest”** as “If the page under test is determined to be relevant to the topic, however, the process moves to block 110, wherein entries are generated for the link table 34 for all outlinks of the page” (Column 10, lines 35-39).

Regarding claim 15, **Chankrabarti** teaches a computer program product comprising:

A) a fifth set of instruction codes for seeding a plurality of seed uniform resource locator strings to start the collaborative focused crawling of the documents (Chakrabarti, Column 5, lines 61-67-Column 6, lines 1-15).

The examiner notes that **Chankrabarti** teaches “**a fifth set of instruction codes for seeding a plurality of seed uniform resource locator strings to start the collaborative focused crawling of the documents**” as “It is to be understood that information pertaining to a “seed” set of Web pages is initially stored in the Web page table 32. The seed set can be gathered from, e.g., the temporary Internet file directories of the employees of a company or from some other group that can be expected to have shared interests... Thus, the seed set does not define a comprehensive, universal set of all topics on the Web, but rather a relatively narrow topic or range of topics that are of interest to the particular source” (Column 5, lines 61-67-Column 6, lines 1-4).

Regarding claim 16, **Chankrabarti** teaches a computer program product comprising:

A) wherein the fourth set of instruction codes further crawls the seed uniform resource locator strings (Chakrabarti, Column 6, lines 61-67-Column 7, lines 1-2, Column 10, lines 44-64).

The examiner notes that **Chankrabarti** teaches “**wherein the fourth set of instruction codes further crawls the seed uniform resource locator strings**” as “starting with the seed set the URL of each page is selected” (Column 6, lines 61-62)

and "the current page is classified to its topics, using the topic analyzer 28 (FIG. 1), and then the page is evaluated for relevancy to the predefined topic at the decision diamond 116...when the page is a "good" page the logic expands the outlinks of the page" (Column 10, lines 45-51).

Regarding claim 17, **Chankrabarti** teaches a computer program product comprising:

A) a sixth set of instruction codes for writing a plurality of resulting uniform resource locator strings obtained by crawling the seed uniform resource locator strings (Column 10, lines 35-43, 51-64).

The examiner notes that **Chankrabarti** teaches "**a sixth set of instruction codes for writing a plurality of resulting uniform resource locator strings obtained by crawling the seed uniform resource locator strings**" as "If the page under test is determined to be relevant to the topic, however, the process moves to block 110, wherein entries are generated for the link table 34 for all outlinks of the page" (Column 10, lines 35-38).

Regarding claim 18, **Chankrabarti** teaches a system comprising:

A) an evaluator that selectively prioritizes the documents to crawl based on a set of rules (Column 8, lines 2-30);
B) a fetcher that fetches prioritized documents from the network (Column 5, lines 40-46);

- C) for each fetched document, a focus engine determines whether the fetched document is relevant to any of the multiple focus topics (Column 2, lines 56-60, Column 3, lines 51-55, Column 4, lines 61-65, Column 10, lines 18-43);
- D) a crawler for crawling the fetched document that matches any of the multiple focus topics (Column 2, lines 56-60, Column 3, lines 51-55, Column 4, lines 61-65, Column 10, lines 35-43); and
- E) wherein the crawler further crawls out-links on the fetched document based on an assumption that if the fetched document is of interest, the out-links are also of interest (Column 2, lines 56-60, Column 3, lines 51-55, Column 4, lines 61-65, Column 10, lines 35-43).

The examiner notes that **Chankrabarti** teaches “**an evaluator that selectively prioritizes the documents to crawl based on a set of rules**” as “the Web page table 32 includes a priority field 42 that represents how often the Web page is to be revisited by the crawler 14” (Column 5, lines 41-42). The examiner further notes that **Chankrabarti** teaches “**for each fetched document, a focus engine determines whether the fetched document is relevant to any of the multiple focus topics**” as “The topic analyzer 28 compares the content of a Web page with a predefined topic or topics and generates a response representative of how relevant the Web page is” (Column 4, lines 61-65), “When the process determines that the page under test is not relevant to the predefined topic” (Column 10, lines 18-19), and “If the page under test is determined to be relevant to the topic” (Column 10, lines 35-36). The examiner further notes that **Chankrabarti** teaches “**a crawler for crawling the fetched document that**

matches any of the multiple focus topics" as "If the page under test is determined to be relevant to the topic, however, the process moves to block 110, wherein entries are generated for the link table 34 for all outlinks of the page" (Column 10, lines 35-39).

The examiner further notes that **Chankrabarti** teaches "**wherein the crawler further crawls out-links on the fetched document based on an assumption that if the fetched document is of interest, the out-links are also of interest**" as "If the page under test is determined to be relevant to the topic, however, the process moves to block 110, wherein entries are generated for the link table 34 for all outlinks of the page" (Column 10, lines 35-39).

Regarding claim 19, **Chankrabarti** teaches a system comprising:

A) a plurality of seed uniform resource locator strings that are used to initiate the collaborative focused crawling of the documents (Chakrabarti, Column 5, lines 61-67- Column 6, lines 1-15).

The examiner notes that **Chankrabarti** teaches "**a plurality of seed uniform resource locator strings that are used to initiate the collaborative focused crawling of the documents**" as "It is to be understood that information pertaining to a "seed" set of Web pages is initially stored in the Web page table 32. The seed set can be gathered from, e.g., the temporary Internet file directories of the employees of a company or from some other group that can be expected to have shared interests... Thus, the seed set does not define a comprehensive, universal set of all

topics on the Web, but rather a relatively narrow topic or range of topics that are of interest to the particular source" (Column 5, lines 61-67-Column 6, lines 1-4).

Regarding claim 20, **Chankrabarti** teaches a system product comprising:

A) wherein the crawler further crawls the seed uniform resource locator strings (Chakrabarti, Column 6, lines 61-67-Column 7, lines 1-2, Column 10, lines 44-64).

The examiner notes that **Chankrabarti** teaches "wherein the crawler further crawls the seed uniform resource locator strings" as "starting with the seed set the URL of each page is selected" (Column 6, lines 61-62) and "the current page is classified to its topics, using the topic analyzer 28 (FIG. 1), and then the page is evaluated for relevancy to the predefined topic at the decision diamond 116...when the page is a "good" page the logic expands the outlinks of the page" (Column 10, lines 45-51).

Claim Rejections - 35 USC § 103

11. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of

the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

12. Claims 10-11 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Chankrabarti et al.** (U.S. Patent 6,418,433) as applied to claims 1-9, and 14-20 and in view of **Heydon et al.** (Article entitled "Mercator: A Scalable, Extensible Web Crawler", dated 06/26/1999).

13. Regarding claim 10, **Chankrabarti** does not explicitly teach a method comprising:

A) the miner allowing a crawling of the resulting resource locator string when the resulting resource locator string matches a plurality of web space rules.

Heydon, however, teaches "**the miner allowing a crawling of the resulting resource locator string when the resulting resource locator string matches a plurality of web space rules**" as "The URL filtering mechanism provides a customizable way to control the set of URLs that are downloaded...The URL filter class has a single crawl method that takes a URL and returns a Boolean value indicating whether or not to crawl that URL" (Page 6, Section: 3.6: URL Filters).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching

Heydon's would have allowed **Chankrabarti's** to provide a scalable and customizable web crawler to fit a specific user's needs, as noted by **Heydon** (Page 2, Section: 2: Related Work).

Regarding claim 11, **Chankrabarti** does not explicitly teach a method comprising:

A) wherein the web space rules comprise domain rules, IP address rules, and prefix rules.

Heydon, however, teaches “**wherein the web space rules comprise domain rules, IP address rules, and prefix rules**” as “Mercator includes a collection of different URL filter subclasses that provide facilities for restricting URLs by domain, prefix, or protocol type” (Page 6, Section: 3.6: URL Filters).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Heydon's** would have allowed **Chankrabarti's** to provide a scalable and customizable web crawler to fit a specific user's needs, as noted by **Heydon** (Page 2, Section: 2: Related Work).

14. Claims 12-13 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Chankrabarti et al.** (U.S. Patent 6,418,433) as applied to claims 1-9, and 14-20 and in view of **Liang** (U.S. PGPUB 2001/0044818).

15. Regarding claim 12, **Chankrabarti** does not explicitly teach a method comprising:

A) the miner disallowing the crawling of the resulting resource locator string when the content of the resulting resource locator string matches a focus topic of the miner.

Heydon, however, teaches “**the miner disallowing the crawling of the resulting resource locator string when the content of the resulting resource locator string matches a focus topic of the miner**” as “Web spider 26 is preferably provided with a copy of the lexicon described above so as to permit it to recognize pornographic material” (Paragraph 62) and “if any page in a website is discovered as comprising pornographic material, all pages “below” that page in the sitemap for the website may be blocked (Paragraph 68).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Liang’s** would have allowed **Chankrabarti’s** to provide a method to allow for web crawlers and spiders to dynamically restrict unwanted and unacceptable material, as noted by **Liang** (Paragraph 3).

Regarding claim 13, **Chankrabarti** does not explicitly teach a method comprising:

A) wherein the miner comprises an unfocus miner that places the resulting uniform resource locator strings that match an unfocus topic in a blacklist, so that the uniform resource locator strings will not be crawled again.

Heydon, however, teaches “**wherein the miner comprises an unfocus miner that places the resulting uniform resource locator strings that match an unfocus**

topic in a blacklist, so that the uniform resource locator strings will not be crawled again" as "web spider 26 determines whether the retrieved web content contains pornographic material. If it does, then in step 908, web spider 26 adds the URL to list 28" (Paragraph 63).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Liang's** would have allowed **Chankrabarti's** to provide a method to allow for web crawlers and spiders to dynamically restrict unwanted and unacceptable material, as noted by **Liang** (Paragraph 3).

Conclusion

16. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

U.S. Patent 6,199,081 issued to **Meyerzon et al.** on 06 March 2001. The subject matter disclosed therein is pertinent to that of claims 1-20 (e.g., methods to specifically crawl targeted subject matter).

U.S. PGPUB 2004/0049514 issued to **Burkov** on 11 March 2004. The subject matter disclosed therein is pertinent to that of claims 1-20 (e.g., methods to specifically crawl targeted subject matter).

U.S. PGPUB 2002/0194161 issued to **McNamee et al.** on 19 December 2002. The subject matter disclosed therein is pertinent to that of claims 1-20 (e.g., methods to specifically crawl targeted subject matter).

U.S. Patent 6,754,873 issued to **Law et al.** on 22 June 2002. The subject matter disclosed therein is pertinent to that of claims 1-20 (e.g., methods to specifically crawl targeted subject matter).

U.S. Patent 7,080,073 issued to **Jiang et al.** on 18 July 2006. The subject matter disclosed therein is pertinent to that of claims 1-20 (e.g., methods to specifically crawl targeted subject matter).

U.S. PGPUB 2006/0277175 issued to **Jiang et al.** on 07 December 2006. The subject matter disclosed therein is pertinent to that of claims 1-20 (e.g., methods to specifically crawl targeted subject matter).

U.S. Patent 6,993,534 issued to **Denesuk et al.** on 31 January 2006. The subject matter disclosed therein is pertinent to that of claims 1-20 (e.g., methods to specifically crawl targeted subject matter).

U.S. Patent 6,295,559 issued to **Emens et al.** on 25 September 2001. The subject matter disclosed therein is pertinent to that of claims 1-20 (e.g., methods to specifically crawl targeted subject matter).

U.S. PGPUB 2002/0032869 issued to **Lamberton et al.** on 14 March 2002. The subject matter disclosed therein is pertinent to that of claims 1-20 (e.g., methods to specifically crawl targeted subject matter).

Contact Information

17. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Mahesh Dwivedi whose telephone number is (571) 272-2731. The examiner can normally be reached on Monday to Friday 8:20 am – 4:40 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Tim Vo can be reached (571) 272-3642. The fax number for the organization where this application or proceeding is assigned is (571) 273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Mahesh Dwivedi

Patent Examiner

Art Unit 2168


December 12, 2006


Leslie Wong
Primary Examiner


TIM VO
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100